Technical Information

Selection of Input Factors and Comparison of Machine Learning Models for Prediction of Dissolved Oxygen in Gyeongang Stream

Min Ji Kim¹ · Seon Jeong Byeon¹ · Kyung Min Kim² · Johng-Hwa Ahn^{1,2,†}

¹Department of Environmental Engineering, Kangwon National University

²Department of Integrated Energy and Infra System, Kangwon National University

(Received December 28, 2020; Revised March 9, 2021; Accepted March 9, 2021)

Objectives : In this study, we select input factors for machine learning models to predict dissolved oxygen (DO) in Gyeongang Stream and compare results of performance evaluation indicators to find the optimal model.

Methods : The water quality data from the specific points of Gyeongang Stream were collected between January 15, 1998 and December 30, 2019. The pretreatment data were divided into train and test data with the ratio of 7:3. We used random forest (RF), artificial neural network (ANN), convolutional neural network (CNN), and gated recurrent unit (GRU) among machine learning. RF and ANN were tested by both random split and time series data, while CNN and GRU conducted the experiment using only time series data. Performance evaluation indicators such as square of the correlation coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE) were used to compare the optimal results for the models.

Results and Discussion : Based on the RF variable importance results and references, water temperature, pH, electrical conductivity, $\text{PO}_4\text{-P}$, $\text{NH}_4\text{-N}$, total phosphorus, suspended solids, and $\text{NO}_3\text{-N}$ were used as input factors. Both RF and ANN performed better with time series data than random split. The model performance was good in order of $\text{RF} > \text{CNN} > \text{GRU} > \text{ANN}$.

Conclusions : The eight input factors (water temperature, pH, electrical conductivity, $\text{PO}_4\text{-P}$, $\text{NH}_4\text{-N}$, total phosphorus, suspended solids, and $\text{NO}_3\text{-N}$) were selected for machine learning models to predict DO in Gyeongang Stream. The best model for DO prediction was the RF model with time series data. Therefore, we suggest that the RF with the eight input factors could be used to predict the DO in streams.

Keywords : Artificial Neural Network, Convolutional Neural Network, Gated Recurrent Unit, Gyeongang Stream, Random Forest

The Korean text of this paper can be translated into multiple languages on the website of <http://jksee.or.kr> through Google Translator.

[†] Corresponding author E-mail: johngghwa@kangwon.ac.kr Tel: 033-250-6357 Fax: 033-259-5550

기술자료

경안천 용존 산소 예측을 위한 입력 인자 선정 및 기계 학습 모델 비교

김민지¹ · 변선정¹ · 김경민² · 안종화^{1,2,*}

¹강원대학교 환경공학과

²강원대학교 신산업개발 T-EMS 융합학과

목적: 본 연구에서는 경안천의 용존 산소(dissolved oxygen, DO) 예측을 위해 기계 학습(machine learning) 모델의 최적 입력 인자를 선정하고 성능 평가 지표 결과를 비교하여 최적 모델을 찾고자 한다.

방법: 경안천 특정 지점의 수질 자료를 연구대상으로 삼아 1998년 1월 15일부터 2019년 12월 30일까지 자료를 수집하고, 전처리한 데이터를 7:3의 비율에 따라 train과 test 자료로 나누어 실험을 진행하였다. 기계 학습 중 랜덤 포레스트(random forest, RF), 인공신경망(artificial neural network, ANN), 합성곱 신경망(convolutional neural network, CNN), 게이트 순환 유닛(gated recurrent unit, GRU) 등을 이용하였다. RF와 ANN은 무작위 추출(random split)과 시계열 자료(time series)로 구분하여 실험하였으며, CNN과 GRU는 시계열 자료만 이용하여 실험을 진행하였다. 모델별 최적의 결과를 비교하기 위해 성능 평가 지표(결정 계수(square of the correlation coefficient, R^2), 평균 제곱근 오차(root mean square error, RMSE), 평균 절대 오차(mean absolute error, MAE))를 사용하였다.

결과 및 토의: RF 기여도 분석 결과와 참고문헌을 통해 최적 입력 인자로 수온, pH, 전기 전도도, $PO_4\text{-P}$, $NH_4\text{-N}$, 총 인, 부유물질, $NO_3\text{-N}$ 등으로 선정하였다. RF와 ANN 모두 무작위 추출보다 시계열 자료의 성능이 더 우수하였다. 시계열 자료를 이용하여 모델 성능을 비교해 보면, RF > CNN > GRU > ANN 순으로 나타났다.

결론: 본 연구에서 경안천의 DO 예측을 위한 기계 학습 모델의 최적 입력 인자로 8개(수온, pH, 전기 전도도, $PO_4\text{-P}$, $NH_4\text{-N}$, 총 인, 부유물질, $NO_3\text{-N}$)를 선택하였다. DO 예측에 가장 우수한 모델은 시계열 자료를 사용한 RF 모델이었다. 따라서 경안천과 같은 하천의 DO를 예측하는 경우 최적 입력 인자 선정 후 시계열 자료를 바탕으로 RF 모델을 이용할 것을 제안한다.

주제어: 인공신경망, 합성곱 신경망, 게이트 순환 유닛, 경안천, 랜덤 포레스트

1. 서론

수질 모니터링은 수자원 관리에서 중요한 역할을 하는데, 수질 예측은 수질 환경 관리에서 필수적이다.^{1,2)} 입력 인자와 출력 인자의 관계를 가정하는 전통적 데이터 처리 방법인 통계 기반 수질 모델은 선형적이며 정규 분포를 따른다.³⁾ 그러나 수질은 이들 사이의 관계가 복잡하고 비선형적이므로 전통적인 데이터 처리 방법으로는 한계가 있다.⁴⁾ 이에 최근에는 다양한 기계 학습(machine learning)과 딥러닝(deep learning) 등을 이용하여 수질을 예측하는 연구가 활발히 이루어지고 있다.

기계 학습은 컴퓨터가 수행해야 할 작업을 데이터로부터 스스로 학습하는 기술로, 인공지능(artificial intelligence, AI) 기술의 하나로써 1980년대부터 개발되기 시작하였다.⁵⁻⁷⁾ 기계 학습의 종류에는 랜덤 포레스트(random forest, RF)와 인공신경망(artificial neural network, ANN) 등이 있는데, RF는

classification and regression trees (CART)를 기반으로 한 결정트리(decision tree)의 확장 개념으로 수많은 독립적인 결정트리를 구축하는 방식이다.⁸⁾ 강수량(rainfall), 수위(water level), 유량(river flow), pump operation을 입력 인자로 사용하여 영국 남해안 지하수 공급원의 탁도(turbidity)를 예측할 때 RF를 사용하기도 하였다.⁹⁾ ANN은 서로 연결된 여러 신경 세포들이 신호를 나누어서 주고받는 신경망 구조와 비슷한 방식으로, 정보를 가지고 있는 데이터가 입력층(input layer)에 입력되면 은닉층(hidden layer)에서 나누어 연산을 수행하며 그 결과를 출력층(output layer)에서 모아 최종적인 결과를 산출하는 방식이다.⁶⁾ 총 인(total phosphorus, TP)을 입력 인자로 사용해 북아메리카의 Okeechobee 호수와 Erie 호수의 TP를 예측하는데, ANN을 사용하기도 하였다.¹⁰⁾

딥러닝은 기계 학습 기술이며 인공신경망을 심화시킨 알고리즘으로, 규모가 큰 신경망을 사용하고 많은 양의 데이터로

학습하는 특성이 있다.^{6,11)} 딥러닝의 종류로는 합성곱 신경망(convolutional neural network, CNN), 게이트 순환 유닛(gated recurrent unit, GRU) 등이 있다.¹²⁾ CNN은 합성곱 층(convolutional layer)과 풀링 층(pooling layer)으로 구성된 ANN으로 합성곱 층에서 가중치 행렬 형태의 필터를 이동시키며 합성곱 연산을 수행하고, 풀링 층을 통해 결과를 추출하는 방식이다.¹³⁾ 수온(water temperature, Temp.), 용존 산소(dissolved oxygen, DO), 전기 전도도(electrical conductivity, EC), 총 유기탄소(total organic carbons), 총 질소(total nitrogen, TN), pH, TP를 입력 인자로 사용하여 대청호의 클로로필 a(chlorophyll-a, Chl-a)를 예측하는데 CNN을 사용하였다.¹⁴⁾ GRU는 가중치 이외의 기억에 관한 추가적인 정보를 셀 상태(cell state)로 저장하여 시계열 패턴의 길이를 조절하는 장단기 메모리(long short-term memory, LSTM)에서 출력 게이트를 제거하여 간소화시킨 형태로¹⁵⁾, 화학적 산소 요구량(chemical oxygen demand, COD)을 입력 인자로 사용하여 상하이 Jinze 저수지의 COD를 예측할 때 GRU를 사용하기도 하였다.¹⁶⁾

DO는 수질 모니터링에 있어 중요한 변수 중 하나로, 수생태계의 상태를 나타내는 중요한 지표이다.^{1,17)} 적절한 범위의 DO는 수생동물의 성장과 발달을 보장하지만, DO 함량이 낮으면 수생동물의 죽음으로 이어질 수도 있다.^{18,19)} 따라서 DO를 예측해 수자원을 관리할 필요가 있다. ANN을 사용하여 DO를 예측한 논문은 다수 있었으나²⁰⁻²⁴⁾ RF²⁾, CNN¹⁾, GRU²⁵⁾ 등을 이용하여 DO를 예측한 논문은 모델별로 한 편 정도밖에 찾을 수 없었다.

경안천은 한강 수계에 속하고, 용인시와 광주시를 관통하면서 서울의 상수원인 팔당호로 유입된다.²⁶⁾ 경안천 유역은 남한강과 북한강보다 매우 적은 수량으로 팔당호로 유입되나

유입 하천 부하가 팔당호 수질에 미치는 영향이 크기 때문에 특별한 관리가 필요하다.²⁷⁾ 이 지역에서 수질 변화를 예측함으로써 수자원을 관리하는 데 도움이 될 수 있다. 하지만 DO를 예측한 논문 중에서 ANN과 비교하면 상대적으로 RF, CNN, GRU 등에 관한 연구가 부족하고, 경안천에서 기계 학습을 이용하여 DO를 예측한 논문은 찾을 수 없었다. 또한 기계 학습 모델의 입력 인자는 결과에 크게 영향을 미치므로 적절한 입력 인자를 선정하는 것이 중요하다.^{17,28)} 따라서 본 연구에서는 경안천의 DO 예측을 위한 입력 인자를 선정하고 기계 학습 중 RF, ANN, CNN, GRU의 성능 평가 지표 결과를 통해 모델을 비교하여 경안천의 DO 예측을 위한 최적 모델을 제시하고자 한다.

2. 연구 방법

2.1. 대상 지역

경안천은 경기도 용인시 호동 문수봉에서 발원하여 용인시를 관통하고 고산천, 직리천, 중대천, 목현천의 지류 하천이 유입된 후 광주시를 관통하면서 곤지암천이 합류된 후 수도권의 상수원인 팔당호로 유입되는 561.12 km²의 면적을 차지하는 국가하천이다.^{26,29)} 경안천 유역은 인구가 밀집해 있으며 많은 하수처리장이 있어 수질 측면에서 팔당호에 영향을 주고 있다.³⁰⁾ 팔당호 유입량에 대한 기여도는 2% 이내이지만, 단위 면적당 오염부하량이 팔당호로 유입되는 한강 수계 중 가장 크므로 특별한 관리가 필요하다.^{29,31)} 총 49.5 km의 경안천 유역 중 본 연구에서는 경안천 A(경기도 광주시 초월읍 무갑리 868)와 경안천 B(경기도 광주시 퇴촌면 정지리 700)를 연구지역으로 선정하였다(Fig. 1). 경안천 A와 B는 광주하수처리장(경기도 광주시 초월읍 경수길 11)의 방류수가 직접

Fig. 1. Location of the study area.

Table 1. Summary of statistical variable for input factors.

	Temp. (°C)	pH (-)	EC (μS/cm)	PO ₄ -P (mg/L)	NH ₄ -N (mg/L)	TP (mg/L)	SS (mg/L)	NO ₃ -N (mg/L)
Maximum	32.9	9.5	822	0.71	8.18	0.76	238.8	8.45
Minimum	0	6.2	33	0	0	0.01	0.2	0.62
Median	16	8	412	0.01	0.17	0.06	8.6	3.28
Mean	15.11	8.01	398.95	0.04	0.6	0.09	12.51	3.44

유입되며, 하수처리장의 방류수가 경안천 유량의 많은 부분을 차지하므로 이에 대한 대책이 필요하다.^{27,29)}

2.2. 자료 수집 및 전처리

경안천 지역의 원본 자료는 물환경정보시스템(<http://water.nier.go.kr>)에서 제공하는 1998년 1월 15일부터 2019년 12월 30일까지 54개의 인자, 총 1,491개의 자료를 사용하였다. 원본 자료 중 누락 값을 제거하는 방향으로 전처리를 진행하여 14개의 인자(부유물질(suspended solids, SS), 생화학적 산소 요구량(biochemical oxygen demand, BOD), 용존 총 인(dissolved total phosphorus, DTP), 용존 총 질소(dissolved total nitrogen, DTN), COD, DO, EC, NH₄-N, NO₃-N, pH, PO₄-P, Temp., TN, TP)로 구성된 총 1,197개의 자료로 축소하였다. 이후 문헌과 RF 기여도 분석(variable importance analysis) 결과를 통해 최적 입력 인자를 설정하였다(Table 1). RF와 ANN은 무작위 추출(random split)과 시계열 자료(time series)로 구분하여 실험하였으며, CNN과 GRU는 시계열 자료만 이용하여 실험하였다. 학습이 용이하도록 시계열 자료만 정규화를 수행하였다. 1,197개의 자료 중 836개(70%)의 자료는 train에 사용하였으며, 나머지 361개(30%)의 자료는 test에 사용하였다.

2.3. 실험 모델

2.3.1. 랜덤 포레스트

RF는 Breiman³²⁾에 의해 제시된 앙상블 학습 방법 모델로 부트스트랩(bootstrap) 방식을 이용하여 다수의 표본을 생성하고, 결정트리를 적용하여 그 결과를 종합하는 방법이다.³³⁾ 결정트리의 node 별로 무작위로 입력 인자를 선택하여 사용함으로써 독립적인 결정트리를 만들어낸다.³⁴⁾(Fig. 2(a)) 이때 결정트리란 주어진 입력 값에 대하여 간단한 조건, 보통 이분법으로 연속적으로 나누어 출력값을 제공하는 방법을 뜻한다.³⁵⁾ 모델의 결과는 결정트리의 개수에 따라 달라지는데, 본 연구에서는 시행 착오법(trial and error method)을 통해 모델별 결정트리를 선정하였다.³⁶⁾

2.3.2. 인공신경망

ANN은 사람의 뉴런(neuron) 세포에서 비롯된 통계적 학습 알고리즘으로, 입력과 출력 자료 사이의 복잡한 비선형 관계

나 패턴을 식별할 수 있으며 train과 test 과정을 바탕으로 출력 값을 추정할 수 있는 유연한 수학 구조다.^{37,38)} ANN은 자료가 모델에 도입되어 입력의 가중 합계 계산이 이루어지는 입력 층, 자료가 처리되는 은닉 층, 결과가 산출되는 출력 층으로 이루어져 있다.³⁹⁾(Fig. 2(b)) 활성화 함수(activation function)는 입력 층으로부터 입력받은 값들을 하나의 값으로 조합하고 이를 출력으로 변환시킨다.⁴⁰⁾ 인공신경망 학습에 있어 일반적인 문제가 있는데 과적합(overfitting)과 과소적합(underfitting) 현상이 그중 하나다. 과적합 문제는 모델이 훈련데이터에만 학습이 집중되어 훈련과정에 사용되지 않은 새로운 데이터에 대해서는 정확도가 낮아지는 문제를 말하고 그 반대를 과소적합이라 한다.⁴¹⁾ 과적합 현상을 해소하려는 방법으로는 earlystop과 dropout이 있다. Earlystop은 과적합이 발생하기 전에 학습을 종료하는 방법이다. Dropout은 은닉 층의 노드를 지정된 비율만큼 무작위로 선택하여 학습에 사용하지 않는 방법이다. 한번 학습을 하고 나면 다음 학습 시에는 다시 노드를 무작위로 선택하여 학습하고 테스트를 할 때는 모든 노드를 사용하여 테스트를 진행한다.⁴²⁾ 또한 은닉 층의 node 개수가 너무 많으면 과적합 문제가 발생할 수 있고, 너무 적으면 과소적합 문제가 발생할 수 있으므로 은닉 층의 node를 결정하는 것이 ANN을 설계할 때 가장 중요하다.⁴³⁾ 하지만 은닉 층의 node 개수를 선정하는 명확한 기준이 없으므로 본 연구에서는 시행 착오법을 이용하여 node 개수를 선정하였다.³⁾

2.3.3. 합성곱 신경망

CNN은 ANN에서 파생되어 나온 기법이다.⁴⁰⁾ 신경망에서 합성곱을 통해 이미지나 배열 형태로 주어진 자료의 특징을 추출하는 방식으로, 최근 다양한 문제 해결에 광범위하게 사용되고 있으며 정확도가 높다.^{44,47)} 은닉 층은 합성곱 층, 풀링 층, 완전히 연결된 층(fully-connected layer)으로 이루어져 있다.¹⁾(Fig. 2(c)) 합성곱 층은 전체 원시 입력의 다양한 영역 또는 학습 가능한 필터(filter)가 있는 중간 특성 맵(map)에서 특성을 자동으로 추출하는 역할을 한다.⁴⁸⁾ 풀링 층은 풀링 창(changing window)의 모든 값을 하나의 값으로 변환하는데, 본 연구에서는 이전 층의 하위 영역에서 최대의 값을 선택하는 최대 풀링(max pooling) 작업으로 입력 층의 크기를 줄였다.^{1,48)} 이러한 작업으로 학습 프로세스의 계산 비용을 줄이고, 과적합

Fig. 2. Structure of (a) random forest, (b-1) artificial neural network with random split data (b-2) artificial neural network with time series data, (c) convolutional neural network, and (d) gated recurrent unit.

문제를 처리한다.⁴⁹⁾ Flatten 층은 다차원 행렬화 되어있는 자료를 1차원 시계열 자료로 변환시켜준다.⁵⁰⁾ 완전히 연결된 층은 이전 층에서 추출된 특징을 출력 층으로 변환하는 역할을 한다.⁵¹⁾

2.3.4. 게이트 순환 유닛

GRU는 기본적인 순환 신경망(recurrent neural network, RNN)의 단점을 보완하기 위해 2014년 처음 제안되었다.⁵¹⁾ GRU는 LSTM의 변형 알고리즘으로, 리셋 게이트(reset gate, r), 업데이트 게이트(update gate, z)로 구성되어 있다.^{52,53)} (Fig. 2(d)) 게이트는 전달할 정보를 결정하는 벡터이며 리셋 게이트는

이전의 메모리와 새로운 입력 값을 어떻게 결합할지 결정하고, 업데이트 게이트는 이전의 메모리를 얼마만큼 전달할지 결정한다.⁵³⁾ 이에 대한 계산 과정은 Eq. 1~4와 같다.³²⁾

$$r_t = \sigma(W_r x_t + W_r h_{t-1}) \quad (1)$$

$$z_t = \sigma(W_z x_t + W_z h_{t-1}) \quad (2)$$

$$\tilde{h}_t = \tanh(W_h x_t + W_h h_{t-1}) \quad (3)$$

$$h_t = (1 - z_t)h_{t-1} + z_t \tilde{h}_t \quad (4)$$

여기서, x_t 는 입력 벡터, h_t 는 출력 벡터(hidden state), r_t 는 리셋 게이트, z_t 는 업데이트 게이트, $\tilde{h}_t$ 는 기억 벡터, W 는 가중치, σ 는 sigmoid 함수를 의미한다.

2.4. 성능 평가 지표

모델의 성능을 평가하기 위해 성능 평가 지표(performance evaluation indicator)로는 결정 계수(square of the correlation coefficient, R^2)(Eq. 5)²³⁾, 평균 제곱근 오차(root mean square error, RMSE)(Eq. 6)²⁴⁾, 평균 절대 오차(mean absolute error, MAE)(Eq. 7)⁵⁴⁾를 사용하였다.

$$R^2 = \left[\frac{n \sum_{i=1}^n y_{obs,i} y_{pre,i} - \left(\sum_{i=1}^n y_{obs,i} \right) \left(\sum_{i=1}^n y_{pre,i} \right)}{\sqrt{\left[n \sum_{i=1}^n y_{obs,i}^2 - \left(\sum_{i=1}^n y_{obs,i} \right)^2 \right] \times \left[n \sum_{i=1}^n y_{pre,i}^2 - \left(\sum_{i=1}^n y_{pre,i} \right)^2 \right]}} \right]^2 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{obs,i} - y_{pre,i})^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{obs,i} - y_{pre,i}| \quad (7)$$

여기서, n 은 자료의 개수, $y_{obs,i}$ 는 관측 값, $y_{pre,i}$ 는 예측 값이다.

3. 결과 및 고찰

3.1. 랜덤 포레스트와 입력 인자

RF 기여도 분석과 참고문헌을 통해 최적 입력 인자를 선정하였다(Fig. 3). 기여도 분석 결과에 의하면 Temp.와 pH가 DO 예측에 중요한 인자인데, 이는 선행논문과 같았다.²²⁾ EC는 기여도 분석 결과 세 번째로 높은 영향을 끼쳤으나 이를 입력

Fig. 3. The variable importance of the input factors for random forest.

인자로 사용한 참고문헌은 많지 않았다.^{1,55)} 하지만 본 연구에서는 EC를 제거 후 실험하였을 때 학습에 실패하여 EC를 입력 인자로 사용하였다. 상대적으로 RF 기여도 분석 결과가 높으며, 참고문헌에서도 입력 인자로 사용한 NH₄-N, NO₃-N, TP, PO₄-P를 입력 인자로 사용하였을 때^{21-23,25)}, 성능 평가 지표 결과가 우수하여 입력 인자로 사용하였다. SS를 입력 인자로 사용한 논문은 찾지 못하였으나 본 연구에서는 이를 입력 인자로 사용하였을 때 성능 평가 지표의 결과가 좋았다. BOD, COD는 DO와 밀접한 관련이 있으나 본 연구에서는 입력 인자로 사용하였을 때 성능 평가 지표 결과가 좋지 않았으며, 이를 입력 인자로 사용하지 않은 논문도 존재하여 입력 인자에서 제외하였다.^{2,20,21)} TN, DTN, DTP 등은 입력 인자로 사용한 문헌이 존재하였으나, 본 연구에서는 입력 인자에 포함되었을 때 성능 평가 지표 결과가 좋지 않아 제외하였다. 따라서 최적 입력 인자는 Temp., pH, EC, PO₄-P, NH₄-N, TP, SS, NO₃-N을 설정하였다. 이는 RF 기여도 분석 결과에서 상위 8개를 선택한 것과 동일하였다(Fig. 3). 이에 모든 모델에서 무작위 추출 입력 인자로 앞서 언급한 8개 인자를 사용하

Table 2. Performance evaluation indicators of the random forest, artificial neural network, convolutional neural network, and gated recurrent unit.

Applied model	Train R ²	Test R ²	Train RMSE	Test RMSE	Train MAE	Test MAE
RF random split	0.94	0.67	0.51	1.33	0.35	0.95
RF time series	0.95	0.71	0.48	1.19	0.44	0.90
ANN random split	0.70	0.59	1.18	1.44	0.82	1.07
ANN time series	0.89	0.55	0.73	1.48	0.35	1.17
CNN	0.95	0.60	0.48	1.38	0.14	1.05
GRU	0.95	0.53	0.48	1.47	0.13	1.15

였고, 시계열 자료 입력 인자는 앞에 언급한 8개 인자에 년(year), 월(month), 일(date)을 추가하여 사용하였다.

RF 무작위 추출의 경우, 결정트리 개수가 30개를 초과하여도 성능 평가 지표의 변동이 생기지 않았기 때문에 최적 결정 트리는 30개로 설정하였다(**Fig. 2(a)**). 이때, Train R^2 0.94, Test R^2 0.67, Train RMSE 0.51, Test RMSE 1.33, Train MAE 0.35, Test MAE 0.95의 결과가 나왔다(**Table 2**). RF 시계열 자료의 경우, 결정트리가 497개보다 작아지면 모든 성능 평가 지표의 값이 하락하는 경향을 보였기 때문에 최적 결정 트리는 497개로 설정하였다(**Fig. 2(a)**). 이 경우, Train R^2 0.95, Test R^2 0.71, Train RMSE 0.48, Test RMSE 1.19, Train MAE 0.33, Test MAE 0.90이 나왔다(**Table 2**). R^2 는 1에 가깝고, RMSE와 MAE는 값이 작을수록 모델의 성능이 우수하므로 시계열 자료가 무작위 추출보다 성능이 더 우수한 것으로 나타났다.^{1,22)}

RF 모델을 사용하여 DO를 예측한 논문은 1편밖에 찾을 수 없었다.²⁾ 이 논문은 년, 월, 일, 탁도, 형광 용해 유기물(fluorescent dissolved organic matter, FDOM), DO, pH, SC, Temp. 등 9개의 입력 인자를 적용하여 하이브리드(hybrid) RF 모델(complete ensemble empirical mode decomposition with adaptive noise random forest, CEEMDAN-RF)과 단일(single)

RF 모델을 사용해 미국 오리건주 Tualatin 강의 DO를 예측하였다. 그 결과 CEEMDAN-RF 모델은 Test RMSE 0.10, Test MAE 0.09가 나왔으며, 단일 RF 모델은 Test RMSE 0.30, Test MAE 0.25가 나왔다(**Table 3**).

3.2. 인공신경망

ANN 무작위 추출은 입력 층, 11개의 은닉 층(node 개수: 1층 200개, 2층 180개, 3~6층 100개, 7~9층 90개, 10~11층 50개), 출력 층(출력 인자: DO)으로 구성하였다(**Fig. 2(b-1)**). 활성화 함수를 sigmoid나 hyperbolic tangent (tanh)로 했을 때 모든 성능 평가 지표가 음수로 나와 최적 활성화 함수를 ReLU로 설정하였다. Batch size 50, epoch 650으로 설정하였으며, 모델의 성능을 평가하고 과적합을 방지하고자 train에 사용된 70%의 자료 중 20%를 검증(validation)에 사용하였다.²⁴⁾ 또한 dropout을 추가했을 때 과적합 현상이 다소 해소되어 비율을 0.2로 하여 적용하였다. 하지만 earlystop을 적용하였을 때는 성능이 저하되어 제외하였다. 그 결과, Train R^2 0.70, Test R^2 0.59, Train RMSE 1.18, Test RMSE 1.44, Train MAE 0.82, Test MAE 1.07이 나왔다(**Table 2**).

ANN 시계열 자료는 입력 층, 11개의 은닉 층(node 개수: 1층 400개, 2층 360개, 3~6층 200개, 7~9층 180개, 10~11층

Table 3. Performance evaluation indicators and input factors used in the previous studies.

Applied model	Input factors	Train R^2	Test R^2	Train RMSE	Test RMSE	Train MAE	Test MAE	Location	Literature
CEEMDAN-RF	date, DO, FDOM, month, pH, SC, temp., turbidity, year	-	-	-	0.10	-	0.09	State of Oregon, Tualatin river	Lu and Ma (2020)
single RF	date, DO, FDOM, month, pH, SC, temp., turbidity, year	-	-	-	0.30	-	0.25	State of Oregon, Tualatin river	Lu and Ma (2020)
ANN	pH, SC, temp.	0.94	0.91	0.40	0.40	-	-	USA, Link and Klamath river	Kisi et al. (2020)
ANN	chl-a, month, $\text{NH}_4\text{-N}$, $\text{NO}_3\text{-N}$, pH, temp.	0.75	0.72	-	-	-	-	Taiwan, Te-Chi reservoir	Kuo et al. (2007)
ANN	ammonia, Cl^- , conductivity, Fe, Mn, nitrates, nitrites, pH, temp., TP	0.97	0.87	-	-	-	-	Serbia, Gruža reservoir	Ranković et al. (2010)
ANN	Cl, COD, K, Na, $\text{NH}_4\text{-N}$, $\text{NO}_3\text{-N}$, pH, PO_4 , T-Alk, T-Hard, TS	0.70	0.76	1.50	1.23	-	-	India, Gomti river	Singh et al. (2009)
ANN	Ca^{2+} , Cl^- , EC, $\text{NH}_4\text{-N}$, $\text{NO}_3\text{-N}$, pH, T-Alk, T-Hard	0.93	0.94	0.43	0.46	-	-	China, Heihe river	Wen et al. (2013)
CNN	EC, ORP, pH, temp.	0.92	0.94	0.63	0.66	0.45	0.52	Greece, Small Prespa lake	Barzegar et al. (2020)
GRU	DO, permanganate index, pH, TP	-	-	-	0.18	-	0.13	China, Qiantang	Li et al. (2019)

100개), 출력 층(출력 인자: DO)으로 구성하였다(**Fig. 2(b-2)**). Batch size 50, epoch 650으로 설정하였으며, 검증에 위한 자료는 train 자료의 20%를 사용하였다. 결과는 Train R^2 0.89, Test R^2 0.55, Train RMSE 0.73, Test RMSE 1.48, Train MAE 0.35, Test MAE 1.17이 나왔다(**Table 2**). 성능 평가 지표를 사용해 무작위 추출과 시계열 자료의 성능을 비교하였을 때, ANN의 최적 모델로 시계열 자료를 선정하였다.

선행논문²⁰⁾에서는 pH, SC, Temp. 등 3개의 인자를 입력 인자로 설정해 미국 Link 강과 Klamath 강의 DO를 예측했을 때, Train R^2 0.94, Test R^2 0.91, Train RMSE 0.40, Test RMSE 0.40이 나왔다(**Table 3**). Kuo의 연구²¹⁾에서는 월, Chl-a, $\text{NH}_4\text{-N}$, $\text{NO}_3\text{-N}$, pH, Temp. 등 6개의 입력 인자를 사용해 대만 Te-Chi 저수지의 DO를 예측하였을 때, Train R^2 0.56, Test R^2 0.52가 나왔다. 다른 선행논문²²⁾에서는 망가니즈(Mn), 아질산염(nitrites), 암모니아(ammonia), 염화 이온(Cl^-), 질산염(nitrates), 철(Fe), pH, SC, Temp., TP 등 10개의 입력 인자를 사용해 세르비아 Gruža 저수지의 DO를 예측했을 때, Train R^2 0.56, Test R^2 0.52가 나왔다. Singh의 연구²³⁾에서는 총 고형물(total solids, TS), 총 경도(total hardness, T-Hard), 총 알칼리도(total alkalinity, T-Alk), BOD, Cl, COD, K, Na, $\text{NH}_4\text{-N}$, $\text{NO}_3\text{-N}$, pH, PO_4 등의 12개 인자를 입력 인자로 사용해 인도 Gomti 강의 DO를 예측했을 때, Train R^2 0.70, Test R^2 0.76, Train RMSE 1.50, Test RMSE 1.23이 나왔다. 또 다른 선행논문²⁴⁾에서는 Ca^{2+} , Cl^- , EC, $\text{NH}_4\text{-N}$, $\text{NO}_3\text{-N}$, pH, T-Alk, T-Hard 등 8개의 인자를 입력 인자로 설정해 중국 Heihe 강의 DO를 예측하였을 때, Train R^2 0.93, Test R^2 0.94, Train RMSE 0.43, Test RMSE 0.46이 나왔다(**Table 3**). 앞서 언급한 선행논문에서는 pH가 중요 인자라고 언급했으며 이는 본 연구 결과와 유사하였다. 하지만 pH 이외의 입력 인자는 문헌별로 달랐으며, 입력 인자에 따라 모델의 결과가 달라졌다. 따라서 pH 이외의 인자를 입력 인자로 설정할 때는 기여도 분석 등을 통해 인자를 선별해 실험을 진행할 필요가 있다.

3.3. 합성곱 신경망

CNN은 입력층, 3개 층(node 개수: 1층 100개, 2층 200개, 3층 100개)으로 이루어진 합성곱 층, max pooling이 적용된 풀링 층, flatten 층, 1개의 층(node: 50)으로 이루어진 완전히 연결된 층, 출력층(출력 인자: DO)으로 구성하였다(**Fig. 2(c)**). 모델의 성능을 평가하기 위해 train에 사용된 70%의 자료 중 10%의 자료를 검증에 사용하였다. CNN의 경우 earlystop, dropout 등을 추가했을 때 성능 평가 지표의 결과가 좋지 않게 나와 사용하지 않았다. 활성화 함수는 sigmoid보다 ReLU를 사용했을 때의 결과가 더 좋아 최적 활성화 함수로 ReLU를 설정하였다. Batch size 40, epoch 350에서 최적의 결과가 나왔는데, 이에 따른 결과는 Train R^2 0.95, Test R^2 0.60, Train

RMSE 0.48, Test RMSE 1.38, Train MAE 0.14, Test MAE 1.05이었다(**Table 2**).

선행논문¹⁾에서는 exponential linear units (ELU)를 활성화 함수로 사용하고, 산화환원전위(oxidation reduction potential, ORP), EC, pH, Temp.를 입력 인자로 사용하여 그리스 Small Prespa 호수의 DO를 예측하였을 때, Train R^2 0.92, Test R^2 0.94, Train RMSE 0.63, Test RMSE 0.66, Train MAE 0.45, Test MAE 0.52가 나왔다(**Table 3**). 선행논문은 본 연구와 달리 최적 활성화 함수로 ELU를 사용하였는데, 모델 특성과 자료 형태 등에 따라 최적 활성화 함수가 달라지므로 기계 학습을 이용해 실험을 진행할 때는 모델에 적합한 활성화 함수를 사용해야 한다.

3.4. 게이트 순환 유닛

GRU는 2개의 은닉 층(node 개수: 1층 120, 2층 300)으로 구성하였다(**Fig. 2(d)**). 검증에 위한 자료는 train 자료의 10%로 설정하였으며, 활성화 함수는 ReLU를 사용하였다. Earlystop, dropout을 추가했을 때 Train R^2 , Test R^2 의 값이 모두 낮게 나와 사용하지 않았다. Batch size 20, epoch 200으로 설정하였으며, 그 결과 Train R^2 0.95, Test R^2 0.53, Train RMSE 0.48, Test RMSE 1.47, Train MAE 0.13, Test MAE 1.15이었다(**Table 2**). GRU를 사용하여 DO를 예측한 논문은 1편정도 찾을 수 있었다.²⁵⁾ 이 논문에서는 활성화 함수로 tanh를 사용하고, 과망간산염 지수(permanganate index), DO, pH, TP 등 4개의 인자를 입력 인자로 설정하여 중국 Qiantang 강의 DO를 예측하였다. 그 결과 Test RMSE 0.18, Test MAE 0.13이었다(**Table 3**).

3.5. 모델별 비교

RF 시계열 자료의 경우 전반적으로 train을 잘하였지만, 최고점과 최저점은 train을 잘하지 못하는 모습을 보였다(**Fig. 4(a-1)**). 반면, CNN, GRU, ANN 시계열 자료는 중후반까지 최고와 최저를 포함해서 train을 잘하는 모습을 보이지만 후반부(CNN: 749~830, GRU: 711~830, ANN: 661~830)에서 train을 하지 못하는 모습을 보였다(**Fig. 4(b-1), (c-1), (d-1)**). 모든 모델에서 전반적인 test 양상은 따라가는 모습을 보이지만, 최고점과 최저점은 test하지 못하는 모습을 보였다(**Fig. 4(a-2), (b-2), (c-2), (d-2)**). 이는 딥러닝 기반 모델 3개가 학습 시에 과적합이 발생한 것으로 보인다. 현재의 결과보다 더 좋은 결과를 얻고자 실험과정에서 dropout과 earlystop을 사용하여 과적합 현상을 해소하고자 하였지만, 이는 해결되기 어려웠다. 결과적으로 성능 평가 지표와 그래프를 통해 네 개의 모델을 비교하였을 때, RF > CNN > GRU > ANN 순으로 성능이 좋았다.

Fig. 4. Observed and predicted DO values in the training and testing periods with (a) random forest, (b) artificial neural network, (c) convolutional neural network, and (d) gated recurrent unit.

4. 결론

경안천의 DO 예측을 위한 기계 학습 모델의 최적 입력 인자로 Temp., pH, EC, $\text{PO}_4\text{-P}$, $\text{NH}_4\text{-N}$, TP, SS, $\text{NO}_3\text{-N}$ 를 선정하였다. 입력 인자 선정할 때 기존 문헌만을 참고하기보다 기여도 분석을 통하여 대상 수계에 따른 적합한 인자 선정을 제안한다. RF와 ANN 모델은 무작위 추출보다 시계열 자료의 결

과가 더 우수하였다. 기계 학습을 이용한 경안천 DO 예측 모델의 성능은 $\text{RF} > \text{CNN} > \text{GRU} > \text{ANN}$ 순으로 나타났다. 즉, 본 연구에서는 시계열 자료를 사용한 RF 모델이 가장 우수한 모델이었으므로 경안천 같은 하천의 DO를 예측할 때 앞서 언급한 입력 인자 8개를 이용한 RF 모델을 사용하는 것을 제안한다.

Acknowledgement

본 논문은 4단계 BK21 사업(강원대학교, 다학제 융합 에너지자원산업 핵심인력 양성사업단)으로 지원된 연구입니다.

References

1. R. Barzegar, M. T. Aalami, J. Adamowski, Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model, *Stoch. Environ. Res. Risk Assess.*, 34, 415-433(2020).
2. H. Lu, X. Ma, Hybrid decision tree-based machine learning models for short-term water quality prediction, *Chemosphere*, 249, 1-12(2020).
3. A. Najah, A. E. Shafie, O. A. Karim, A. H. E. Shafie, Application of artificial neural networks for water quality prediction, *Neural. Comput. Appl.*, 22, 187-201(2013).
4. M. Salari, E. S. Shahid, S. H. Afzali, M. Ehteshami, G. O. Conti, Z. Derakhshan, S. N. Sheibani, Quality assessment and artificial neural networks modeling for characterization of chemical and physical parameters of potable water, *Food Chem. Toxicol.*, 118, 212-219(2018).
5. I. J. Kim, Deep learning: a new trend in machine learning, *KICS*, 31(11), 52-57(2014).
6. C. Y. Lee, S. M. Kim, Y. S. Choi, Case analysis for introduction of machine learning technology to the mining industry, *KSRM*, 29(1), 1-11(2019).
7. D. E. Rumelhart, J. L. McClelland, *Learning Internal Representations by Error Propagation*, MIT Press, Cambridge, USA, pp. 18-362(1987).
8. E. N. Jang, J. H. Im, S. H. Ha, S. G. Lee, Y. G. Park, Estimation of water quality index for coastal areas in Korea using GOCI satellite data based on machine learning approaches, *Korean J. Remote. Sens.*, 32(3), 221-234(2016).
9. M. Stevenson, C. Bravo, Advanced turbidity prediction for operational water supply planning, *Decis. Support Syst.*, 119, 72-84(2019).
10. C. Ding, F. Pu, C. Li, X. Xu, T. Zou, X. Li, Combining artificial neural networks with causal inference for total phosphorus concentration estimation and sensitive spectral bands exploration using MODIS, *Water*, 12(9), 1-15(2020).
11. H. I. Gu, Artificial intelligence and deep learning trends, *Korean. Inst. Electr. Eng.*, 67(7), 7-12(2018).
12. H. J. Chun, H. S. Yang, A study on prediction of housing price using deep learning, *J. Res. Environ. Inst. Korea*, 17(2), 37-49(2019).
13. D. C. Woo, H. S. Moon, S. B. Kwon, Y. H. Cho, A deep learning application for automated feature extraction in transaction-based machine learning, *J. Inf. Technol. Serv.*, 18(2), 143-159(2019).
14. J. H. Choi, J. Y. Kim, J. H. Won, O. G. Min, Modelling chlorophyll-a concentration using deep neural networks considering extreme data imbalance and skewness, in *Proceedings of the 2019 21st International Conference on Advanced Communication Technology (ICACT)*, PyeongChang, Korea, pp. 631-634(2019).
15. S. E. Moon, S. B. Jang, J. H. Lee, J. S. Lee, Machine learning and deep learning technology trends, *J. Korean Inst. Commun. Inf. Sci.*, 33(10), 49-56(2016).
16. J. Tu, X. Yang, C. Chen, S. Gao, J. Wang, C. Sun, Water quality prediction model based on GRU hybrid network, in *Proceedings of the 2019 Chinese Automation Congress (CAC)*, Hangzhou, China, pp. 1893-1898(2019).
17. E. Olyaei, H. Z. Abyaneh, A. D. Mehr, A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River, *Geosci. Front.*, 8(3), 517-527(2017).
18. Y. Liu, Q. Zhang, L. Song, Y. Chen, Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction, *Comput. Electron. Agric.*, 165, 1-11(2019).
19. B. M. Gillanders, T. S. Elsdon, I. A. Halliday, G. P. Jenkins, J. B. Robins, F. J. Valesini, Potential effects of climate change on Australian estuaries and fish utilising estuaries: a review, *Mar. Freshw. Res.*, 62(9), 1115-1131(2011).
20. O. Kisi, M. Alizamir, A. D. Gorgij, Dissolved oxygen prediction using a new ensemble method, *Environ. Sci. Pollut. Res.*, 27, 9589-9603(2020).
21. J. T. Kuo, M. H. Hsieh, W. S. Lung, N. She, Using artificial neural network for reservoir eutrophication prediction, *Ecol. Model.*, 200(1-2), 171-177(2007).
22. V. Ranković, J. Radulović, I. Radojević, A. Ostojić, L. Čomić, Neural network modeling of dissolved oxygen in the Gruža Reservoir, Serbia, *Ecol. Model.*, 221(8), 1239-1244(2010).
23. K. P. Singh, A. Basant, A. Malik, G. Jain, Artificial neural network modeling of the river water quality-a case study, *Ecol. Model.*, 220(6), 888-895(2009).
24. X. Wen, J. Fang, M. Diao, C. Zhang, Artificial neural network modeling of dissolved oxygen in the Heihe River, Northwestern China, *Environ. Monit. Assess.*, 185, 4361-4371(2013).
25. L. Li, P. Jiang, H. Xu, G. Lin, D. Guo, H. Wu, Water quality prediction based on recurrent neural network and improved evidence theory: a case study of Qiantang River, China, *Environ. Sci. Pollut. Res.*, 26, 19879-19896(2019).
26. Ministry of Environment, Water Environment Management Plan a Medium Influence Areas in Kyeongan Stream, Ministry of Environment, Sejongsi, Korea(2019).
27. J. H. Jang, C. G. Yoon, K. W. Jung, S. B. Lee, Characteristics of pollution loading from Kyongan Stream watershed by BASINS/SWAT, *Korean J. Limnol.*, 42(2), 200-211(2009).
28. S. Palani, S. Y. Liong, P. Tkachik, An ANN application for water quality forecasting, *Mar. Poll. Bull.*, 56(9), 1586-1597(2008).
29. J. H. Choi, J. H. Ha, S. S. Park, Estimation of the effect of water quality management policy in Paldang Lake, *J. Korean Soc. Environ. Eng.*, 30(12), 1225-1230(2008).
30. S. J. Hwang, K. H. Kim, C. H. Park, W. B. Seo, B. G.

- Choi, H. S. Eum, M. H. Park, H. R. Noh, Y. B. Sim, J. K. Shin, Hydro-meteorological effects on water quality variability in Paldang Reservoir, confluent area of the South-Han River-North-Han River-Gyeong Stream, Korea, *Korean J. Ecol. Environ.*, 49(4), 354-374(2016).
31. C. G. Kim, N. W. Kim, J. E. Lee, Assessing the effect of upstream dam outflows and river water uses on the inflows to the Paldang Dam, *J. Korea Water Resour. Assoc.*, 47(11), 1017-1026(2014).
32. L. Breiman, Bagging predictors, *Mach. Learn.*, 24, 123-140 (1996).
33. J. D. Suh, Foreign exchange rate forecasting using the GARCH extended random forest model, *J. Ind. Econ. bus.*, 29(5), 1607-1628(2016).
34. L. Breiman, Random forests, *Mach. Learn.*, 45, 5-32(2001).
35. G. G. Lee, E. H. Lee, S. W. Kim, K. M. Kim, D. J. Kim, Modeling of flow-accelerated corrosion using machine learning: comparison between random forest and non-linear regression, *Corros. Sci. Tech.*, 18(2), 61-71(2019).
36. J. K. Kim, K. B. Lee, S. G. Hong, ECG-based biometric authentication using random forest, *J. Inst. Korean. Electr. Inf. Eng.*, 54(6), 100-105(2017).
37. J. W. Lee, H. J. Kim, M. K. Kim, Design of short-term load forecasting based on ANN using bigdata, *Korean Inst. Electr. Eng.*, 69(6), 792-799(2020).
38. D. Antanasijević, V. Pocajt, D. Povrenović, A. P. Grujić, M. Ristić, Modelling of dissolved oxygen content using artificial neural networks: Danube River, North Serbia, case study, *Environ. Sci. Pollut. Res.*, 20, 9006-9013(2013).
39. M. J. Diamantopoulou, V. Z. Antonopoulos, D. M. Papamichail, Cascade correlation artificial neural networks for estimating missing monthly values of water quality parameters in rivers, *Water Resour. Manag.*, 21, 649-662 (2007).
40. S. S. Kim, J. W. Kim, Customer behavior prediction of binary classification model using unstructured information and convolution neural network: the case of online storefront, *J. Intell. Inf. Syst.*, 24(2), 221-241(2018).
41. S. H. Choi, M. S. Do, Prediction of asphalt pavement service life using deep learning, *Int. J. Highw. Eng.*, 20(2), 57-65(2018).
42. D. J. Park, B. W. Kim, Y. S. Jeong, C. W. Ahn, Deep neural network based prediction of daily spectators for Korean baseball league : focused on Gwangju-KIA champions field, *Smart Media J.*, 7(1), 16-23(2018).
43. W. K. Yeo, Y. M. Seo, S. Y. Lee, H. K. Jee, Study on water stage prediction using hybrid model of artificial neural network and genetic algorithm, *J. Korea Water Resour. Assoc.*, 43(8), 721-731(2010).
44. W. W. Nam, B. W. Kim, Deep learning based depression classification using environmental factor selection, *Trans. Korean. Inst. Electr. Eng.*, 69(7), 1102-1110(2020).
45. A. Hejnal, F. Rentzsch, Neural nets, *Curr. Biol.*, 25(18), 782-789(2015).
46. B. Jafrasteh, N. Fathianpour, A hybrid simultaneous perturbation artificial bee colony and back-propagation algorithm for training a local linear radial basis neural network on ore grade estimation, *Neurocomputing*, 235(26), 217-227(2017).
47. X. Ta, Y. Wei, Research on a dissolved oxygen prediction method for recirculating aquaculture systems based on a convolution neural network, *Comput. Electron. Agr.*, 145, 302-310(2018).
48. R. Zuo, Y. Xiong, J. Wang, E. J. M. Carranza, Deep learning and its application in geochemical mapping, *Earth Sci. Rev.*, 192, 1-14(2019).
49. E. Hoseinzade, S. Haratizadeh, CNNpred: CNN-based stock market prediction using a diverse set of variables, *Expert Syst. Appl.*, 129(1), 273-285(2019).
50. H. C. Jung, Y. G. Sun, D. G. Lee, S. H. Kim, Y. M. Hwang, I. S. Sim, S. K. Oh, S. H. Song, J. Y. Kim, Prediction for energy demand using 1D-CNN and bidirectional LSTM in internet of energy, *J. Inst. Korean. Electr. Electron. Eng.*, 23(1), 134-142(2019).
51. K. H. Cho, B. V. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Doha, Qatar, pp. 1724-1734(2014).
52. D. Zhang, G. Lindholm, H. Ratnaweera, Use long short-term memory to enhance internet of things for combined sewer overflow monitoring, *J. Hydrol.*, 556, 409-418(2018).
53. M. H. Lee, Y. R. Yoon, H. J. Moon, Performance evaluation of an indoor temperature forecasting model based on GRU for floor heating system operation, *J. KSLES*, 27(3), 272-282(2020).
54. M. Ay, O. Kisi, Modeling of dissolved oxygen concentration using different neural network techniques in foundation creek, El Paso county, Colorado, *J. Environ. Eng.*, 138(6), 654-662(2012).
55. S. M. Lee, Y. G. Sun, J. Y. Lee, D. G. Lee, E. I. Cho, D. H. Park, Y. B. Kim, I. S. Sim, J. Y. Kim, Short-term power consumption forecasting based on IoT power meter with LSTM and GRU deep learning, *J. Inst. Internet Broadcast. Commun.*, 19(5), 79-85(2019).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors and Contribution Statement —————

Min Ji Kim

Department of Environmental Engineering, Kangwon National University, Undergraduate Student, ORCID® 0000-0002-9066-8762: Conceptualization, Data analysis, Validation, Visualization, Writing - original draft.

Seon Jeong Byeon

Department of Environmental Engineering, Kangwon National University, Undergraduate Student, ORCID® 0000-0002-4579-8867: Conceptualization, Data curation, Data analysis, Visualization, Writing - original draft.

Kyung Min Kim

Department of Integrated Energy and Infra System, Kangwon National University, Graduate Student, ORCID® 0000-0002-7348-9367: Methodology, Resources.

Johng-Hwa Ahn

Department of Environmental Engineering, Kangwon National University, Department of Integrated Energy and Infra System, Kangwon National University, Professor, ORCID® 0000-0001-5745-5338: Supervision, Writing - review and editing.